DOI: 10.15849/IJASCA.240330.10

Int. J. Advance Soft Compu. Appl, Vol. 16, No. 1, March 2024
Print ISSN: 2710-1274, Online ISSN: 2074-8523

Time Series Forecasting Model for the Stock Market using LSTM and SVR

Qusai Q. Abuein [1], Mohammed Q. Shatnawi [2], Emran Y. Aljawarneh [3], and Ahmed Manasrah [4]

[1,2] Jordan University of Science and Technology, Department of Computer Information Systems, Irbid, Jordan
E-mail: {qabuein, mshatnawi}@just.edu.jo
[3] Jordan University of Science and Technology, Department of Computer Information Systems, Irbid, Jordan
e-mail: eyaljawarneh19@cit.just.edu.jo
[4] Yarmouk University, Department of Computer Science, Irbid, Jordan
e-mail: Ahmad.a@yu.edu.jo
[4] Higher Collages of Technology, Department of Information Science, Sharjah, UAE
E-mail: amanasrah@hct.ac.ae

Abstract

Time series data prediction is an essential area of research in finance, and economics, among others. It involves analyzing and modeling data collected over time to make future predictions or forecast future trends. With the increasing availability of historical data and advancements in machine learning and deep learning techniques, time series data prediction has become an increasingly popular research topic in recent years. In this paper, we investigate the application of machine and deep learning models to time-series data for predicting the optimal time for trading stocks and options. Time-series data is defined as a collection of historical data points ordered by time, commonly used in predicting stock prices, stock indexes, and cryptocurrency prices. The study uses a publicly available dataset of Japanese stocks and options to train and test Support Vector Regression (SVR) and Long Short-Term Memory (LSTM) models. The goal of the research is to improve trading strategies by identifying the best times to buy and sell assets based on predictive models. The performance of the model are compared using three accuracy measurements: Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). The study has shown that The LSTM with dropout technique provided the best possible results with MSE 0.000124763, RMSE 0.011169727, and MAE 0.009058733.

Keywords: *Long Short-Term Memory, Stock Market, Support Vector Machine, Time Series, Time Series Forecasting.*

1 Introduction

Within the financial stock market, a time series refers to a sequential record of a particular share's price. The ability to anticipate the direction of financial time series trends is

Received 10 September 2023; Accepted 26 February 2024

especially critical for successful investing. Time series analysis [1] has numerous practical applications across a wide range of fields such as finance, economics, engineering, weather forecasting, and more. In finance, time series analysis is used to forecast stock prices, currency exchange rates, and commodity prices. In economics, it is used to model and forecast economic growth, inflation, and unemployment. In engineering, it is used to monitor and predict the performance of mechanical systems, such as engines and turbines. In weather forecasting, time series analysis is used to predict future weather patterns based on past data. Overall, time series analysis provides valuable insights into the past and future behavior of complex systems and helps decision-makers make informed decisions.

Financial time series typically exhibit characteristics such as volatility clustering, non-stationarity, and dependence on past values. These properties make forecasting future prices a challenging task. Time series forecasting is particularly relevant in the field of finance, where stock prices and other financial metrics are often the subject of analysis. Stock prices, for example, can exhibit complex patterns over time, including trends, seasonality, and other cyclical fluctuations, which can make it challenging to accurately predict future prices. In addition, stock prices are influenced by a wide range of factors, including economic indicators, company financials, and market sentiment, which can change rapidly and unpredictably. As a result, time series forecasting models must account for these factors and incorporate uncertainty to provide useful insights and predictions for investors and financial analysts. [2, 3]. A statistical or machine learning model (i.e. a time series forecasting model) is required to analyze the patterns and trends observed in historical data, such as seasonality, trend, and cyclical fluctuations, and uses this information to make predictions about future values [4, 3]. Time series forecasting models may be used to forecast a wide range of variables, including stock prices, weather patterns, economic indicators, and cryptocurrency price prediction [5].

When it comes to making investment decisions, there are various approaches that investors can use to predict the future performance of an asset. Two traditional methods are technical analysis and fundamental analysis [6]. Technical analysis and fundamental analysis are two approaches investors use to make decisions about buying and selling assets. Technical analysis involves analyzing past market data to identify patterns and trends, while fundamental analysis involves examining the underlying financial and economic factors that drive an asset's performance. Investors often use a combination of both techniques to inform their decisions. With advancements in machine learning and deep learning techniques, researchers have developed predictive models that can forecast future stock prices accurately. These models have become popular among investors and traders as they can identify profitable trading opportunities and mitigate risks. The lack of knowledge and uncertainty about when to buy or sell stocks is a common problem faced by many individuals and organizations in the financial field. This uncertainty can make the process of generating profits challenging and can result in significant losses. However, with the advancements in predictive models and machine learning algorithms, investors can now make informed decisions and minimize the risks associated with stock trading. By leveraging historical observations and patterns, these models can forecast future stock prices accurately, providing investors with the necessary knowledge to make the right decisions at the right time. This knowledge can be the key to unlocking profitable investment opportunities and avoiding costly mistakes [7–9].

Financial analysts commonly use machine learning techniques for stock price prediction in the era of big data [7]. These techniques can improve the accuracy of predictions and help investors make informed decisions. However, stock price prediction remains one of the most challenging tasks in financial forecasting due to the complex nature of the stock market [8]. The research in economics, machine learning, and other fields has revolutionized stock price forecasting. Profitability is a top concern for most companies, owners of capital, and startups [9]. Knowing the best time to invest, understanding price stability, and tracking changes in prices over time can help ensure profit. Therefore, the objective of this research is to develop an accurate predictive model that can assist investors in making informed decisions and generating profits through answering the following questions:

1. How can we prepare a suitable dataset for building a robust time series forecasting model?
2. What is the best time series forecasting model?
3. What is the correct time for trading based on the best time series forecasting model?

This study provides valuable insights into the potential of machine and deep learning models for improving the accuracy and profitability of stock and options trading through the following contributions:

1. A thorough investigation of the application of machine and deep learning models to time-series data for predicting the optimal time for trading stocks and options.
2. A comparative study of different models, including traditional statistical models, machine learning models, and deep learning models.
3. Evaluation of the performance of the models using real-world financial data and comparison of their effectiveness in predicting the optimal trading time.
4. Identification of the most promising approaches for future research and development in this field.

This paper began with an introduction to the importance of stock price prediction in the financial field and the challenges associated with it. A literature review will be conducted in Section 2 to examine the existing research on stock price prediction and identify gaps in the current knowledge. The methodology section 3 will detail the data collection process and the machine learning algorithms used for predictive modeling. The results section 4 will present the findings of the study, including the accuracy of the predictive model and the significance of the variables used. Finally, the discussion section 5 will interpret the results and provide insights into the practical implications of the research. The paper will conclude with recommendations for future research to enhance the accuracy of stock price prediction models and improve their applicability to real-world investment scenarios.

2 Related Work

In the context of financial time series analysis, various approaches have been taken to predict prices or trends. These approaches can be broadly classified into four categories: traditional time series analysis methods, machine learning methods, deep learning methods, and combinations of these methods. Traditional time series analysis methods usually involve statistical models that try to fit a time series data to a certain pattern or trend [10, 11]. Machine learning methods, on the other hand, involve the use of algorithms that learn patterns and relationships from the data and make predictions based on these patterns [12, 13]. Deep learning methods, a subfield of machine learning, use neural networks to learn hierarchical representations of the data, enabling them to capture complex patterns and relationships that may be difficult to discern using traditional methods [14–16]. Combining these methods often leads to more accurate predictions, as the strengths of each approach can complement each other [17, 18].

K.-j. Kim [19] investigated the applicability of the Support Vector Machine (SVM), back-propagation neural network (BPN), and case-based reasoning (CBR) to predict future direction of stock price index. However, the authors found that the prediction performance of SVMs in financial time series forecasting is sensitive to the values of its parameters. Therefore, it is crucial to find the optimal value of these parameters to improve the accuracy of the predictions. M. K. Okasha, [2] investigated the effectiveness of support vector machines (SVM) in financial forecasting by comparing it with autoregressive integrated moving average (ARIMA) and artificial neural network (ANN) models. The study used Al-Quds Index data from the Palestinian Stock Exchange Market and aimed to forecast two-month future points. SVM outperformed both ARIMA and ANN models in terms of accuracy and precision, indicating that SVMs are a promising approach to financial time series forecasting. The study highlights the potential of SVM as an effective tool for financial forecasting and provides insights into its application in the context of the Palestinian Stock Exchange Market. Y. Ensafi et al. [20] apply various forecasting models including classical time-series techniques such as Seasonal Autoregressive Integrated Moving Average (SARIMA) and Triple Exponential Smoothing, and more advanced methods like Prophet, Long Short-Term Memory (LSTM), and Convolutional Neural Network (CNN). The performances of these models are compared using different accuracy metrics such as Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE). The findings reveal that the Stacked LSTM method outperforms the other models, while the Prophet and CNN models also perform well.

W.-H. Chen [21] tried to construct a prediction model for forecasting the six major Asian stock market indices using Support Vector Machines (SVMs) and Back Propagation (BP) neural networks. SVM and BP performed better than AR (1) in the deviation performance criteria, but the direction criteria AR (1) model was better than SVM and BP. In the study conducted by Dutta et al. [22], the prediction of stock prices was attempted by utilizing machine learning and sentiment analysis techniques. A hybrid model, comprising time-series and sentiment analysis models, was implemented, and the model (LSTM-VDR) was constructed using Long-Short Term Memory (LSTM) for time-series analysis and Valence Aware Dictionary and Sentiment Reasoner (VADER) for sentiment analysis. The accuracy of the model was compared with that of different combination models, and it was determined that a high accuracy value of 77.496% was achieved by LSTM-VDR. An attempt was made by Tsai and Wang [23] to develop a model that improves the accuracy of stock price forecasting by combining Artificial Neural Network (ANN) and Decision Tree (DT). Their findings indicate that the DT+ANN model outperforms other models, with a 77% accuracy in forecasting electronic industry stock prices. In another study by Siami et al. [24], a comparison was made between Auto-Regressive Integrated Moving Average (ARIMA) and Long-Short Term Memory (LSTM) as forecasting methods to determine the best prediction accuracy. The results of the study showed that the LSTM model performed better than the ARIMA model.

Several predictive models based on machine learning and deep learning were proposed for stock price prediction. For instance, Mehtab and Sen [25] developed multiple models for predicting the National Stock Exchange Fifty (NIFTY 50) stock price movements in the National Stock Exchange of India (NSE). These models were constructed using eight regression and eight classification methods for stock price and movement prediction on weekly forecast horizons. In addition, three Convolutional Neural Network (CNN) models were designed using univariate and multivariate approaches, with varying input data sizes and network configurations. These models were further augmented to improve performance. Time series have been widely applied in different fields, including cryptocurrency price prediction. Several studies, such as [26–28], employed different time series analyses, machine learning, and deep learning models to predict bitcoin prices, and the ARIMA model was found to be the best. Another study by Sin et al. [29] aimed to investigate the influence of bitcoin features on the next day's price change and found that the GASEN model was effective in performing the classification task.

Table 1 summarizes the main characteristics of four different approaches used for financial time series analysis: traditional time series analysis methods, machine learning methods, deep learning methods, and combinations of these methods. It highlights the strengths and weaknesses of each approach and how they complement each other. Where, Table 2 compares the accuracy and performance of different predictive models used for stock price forecasting. It includes models based on traditional time series analysis, machine learning, and deep learning techniques, as well as hybrid models that combine different approaches. The table provides insights into the strengths and limitations of each model and how they can be used for stock price forecasting.

Table 1. Summary of the different approaches to financial time series analysis mentioned and their strengths and weaknesses

Method	Description	Strengths	Weakness
Traditional time series analysis	Involves statistical models that try to fit time series data to certain pattern or trend.	Easy to interpret, computationally efficient.	Limited ability to capture complex patterns.
Machine Learning	Involves algorithms that learn patterns and relationships from data and make predictions based on these patterns.	Can capture complex patterns, adaptable to different types of data.	Prone to overfitting, difficult to interpret.
Deep Learning	A subfield of machine learning that uses neural networks to learn hierarchical representations of	Can capture complex patterns, adaptable to different types of data.	Requires large amounts of data and computational resources, difficult to interpret.

	data, enabling them to capture complex patterns and relationships that may be difficult to discern using traditional methods.		
Combination of Methods	Combining traditional time series analysis, machine learning, and deep learning methods often leads to more accurate predictions as the strengths of each approach can complement each other.	Can capture complex patterns and trends from different angles.	Can be computationally expensive, difficult to interpret.

3 Problem Formulations or Methodology

This study aims to forecast time-series data for a stock market using Support Vector Regression (SVR) and Long Short-Term Memory (LSTM) models. The methodology has three phases, Data preprocessing, stationary test, fitting, and splitting. Data preprocessing involves cleaning and transforming the data to prepare it for analysis. Stationary Test is a statistical method used to check if the time series data is stationary or not. If the data is non-stationary, it is transformed into a stationary time series to enable forecasting. Fitting and Splitting involve training and testing the models using the prepared data. The models are fitted to the training data and evaluated on the testing data to select the best-performing model for forecasting.

Table 2. Summary of Specific prediction models and their respective performances

Model	Description	Performance
Support Vector Machine (SVM)	A machine learning method that finds the optimal hyperplane to separate data into different classes.	Outperformed ARIMA and artificial neural network (ANN) models in financial forecasting.
Long-Short Term Memory (LSTM)	A deep learning method that uses recurrent neural networks to model sequential data, such as time series.	Outperformed Auto-Regressive Integrated Moving Average (ARIMA) in stock price prediction.
Decision Tree (DT) + Artificial Neural Network (ANN)	A combination of machine learning and deep learning methods that improves the accuracy of stock price forecasting.	Achieved 77% accuracy in forecasting electronic industry stock prices.

LSTM-VDR	A hybrid model that combines time-series and sentiment analysis models, using LSTM for time-series analysis and Valence Aware Dictionary and Sentiment Reasoner (VADER) for sentiment analysis.	Achieved high accuracy value of 77.496% in stock price prediction.
ARIMA	A traditional time series analysis method that models time series data as a combination of autoregressive, integrated, and moving average components.	Found to be the best method for predicting bitcoin prices.

3.1 Exploratory Data Analysis (EDA)

In this research, we used a dataset created by the Japan Exchange Group, a holding company that operates one of the largest stock exchanges in the world. The dataset was collected from the Kaggle website for data scientists and machine learning practitioners. It contains historical data for a variety of Japanese stocks and options from 2017 to 2022, with the goal of predicting future returns. The dataset consists of multiple files located in different folders, but we only used three files in two folders: train and test. The stock prices and secondary stock prices files contain the features used in this research, except for the target feature which is time. To obtain this feature, we used the financials file to extract the DisclosedUnixTime feature. Once we merged these files, we obtained the dataset we needed, which consisted of 39,944 data points and 6 features. The attributes used in this research are displayed in Table 3. The dataset showed some non-seasonality in its patterns, so we converted it to show seasonality in its patterns, as displayed in the stationary test section.

3.2 Dataset

The time-series data is considered numerical data points. Financial data is an example of this type. It consists of numeric variables. Time series data is usually fractional numbers (also known as decimal numbers), which takes any value between two numbers, known as continuous data. This data contains historical data for a variety of Japanese stocks and options. The data used in this study were obtained from the Kaggle website for data scientists and machine learning practitioners. The dataset was created by the Japan Exchange Group, which operates one of the largest stock exchanges in the world. It contains historic data for various Japanese stocks and options starting from 4/1/2017 and ending in 27/5/2022.

The dataset consists of multiple folders with several CSV files. To generate a suitable time-series dataset, we used 3 files existed in the original dataset source which are stock prices, secondary stock prices, and financials CSV files to train our model. The dataset included six variables, including Open, High, Low, Close, Volume, and DisclosedUnixTime (timestamp), with nearly 40,000 data points. These variables are displayed in Table 3.

Table 3. Dataset Features

Features	Description
Open	First traded price
High	Highest traded price
Low	Lowest traded price
Close	Last traded price
Volume	Number of traded price
DisclosedUnixTime	Unix time of the datetime on which the document disclosed.

3.4 Preprocessing

Checking Null values, missing values, and duplicated values are the first step in preprocessing the dataset. The dataset we have built for predicting the future returns of stocks. But, we need to make it suitable for predicting the appropriate time for trading. So, we used the 3 files mentioned above to generate a time-series data that is suitable for predicting the appropriate time for trading by concatenating the stock_price.csv with the secondary_stock_prices.csv under the name of Stock_Prices.csv. Moreover, we merged this file with the financials.csv that contains the target feature we need which is the DisclosedUnixTime (timestamp). All these steps and others presented in the Fig.1.

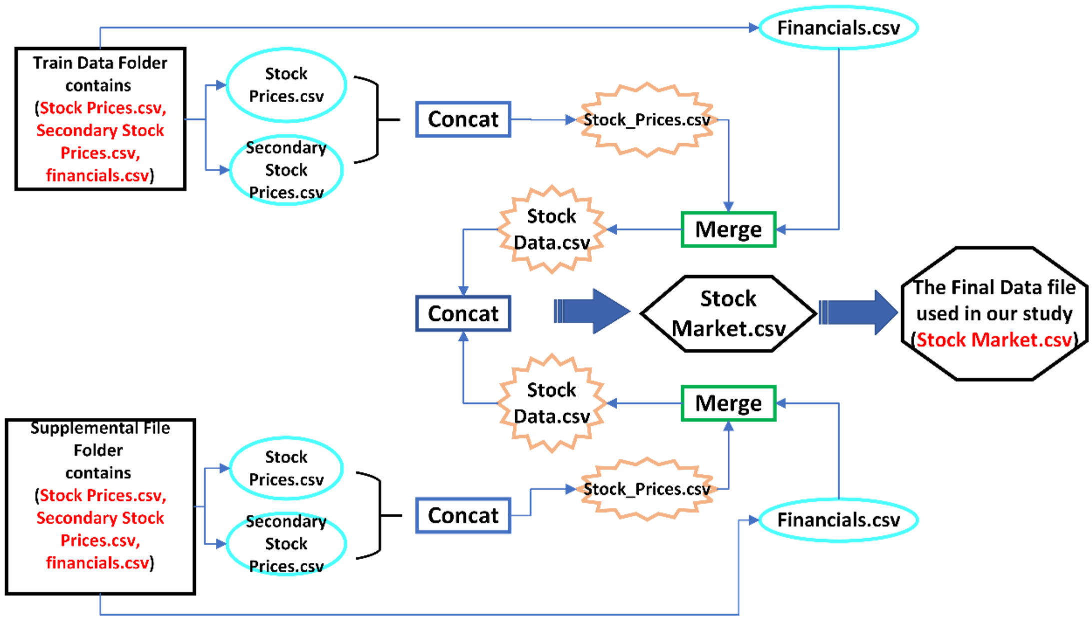

Fig 1. Preprocessing steps

3.4.1 Stationary test

Each time series dataset should be tested to determine if it exhibits stationarity, meaning that its statistical properties such as mean and variance are constant over time. Summary statistics are one of several methods used to assess the non-stationarity of a time series. This involves splitting the time series into two or more partitions and comparing the mean and variance of each group (Table 4). If the means and variances differ significantly between the groups, the time series is likely non-stationary. [30].

Table 4. Stationary and non-stationary test

	Original Data	Transformed Data
Mean 1	216967974.085951	2.237308
Mean 2	234414151.478916	2.192503
Variance 1	282081475428882720.000000	0.148831
Variance 2	329597435678650944.000000	0.166345

A big difference between the mean and variance as shown in the original data indicates that the data is non-stationary. To make it stationary, we applied the natural logarithm to the data. Fig. 2 and Fig. 3 present the behavior of the data before and after the stationary test. Fig. 3 shows the suitable format that is accepted by the model.

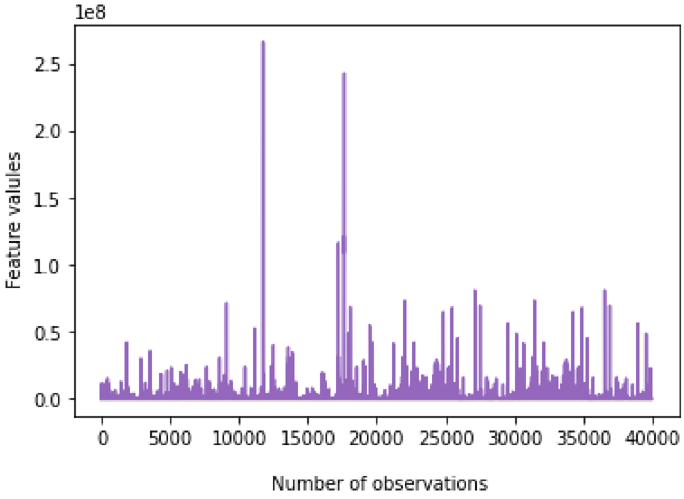

Fig 2. Dataset Distribution

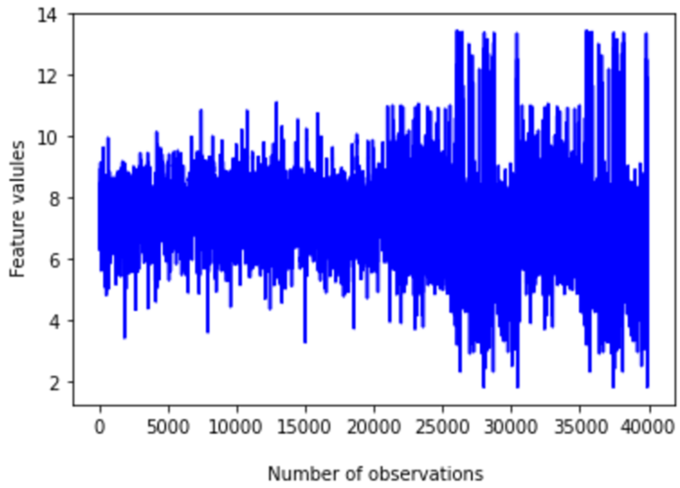

Fig 3. Transformed Dataset

3.5 Fitting the SVM Data

SVM is a supervised machine learning algorithm used for classification, regression, and anomaly detection. In this study, we used Support Vector Regression (SVR) as a type of Support Vector Machine algorithm for regression analysis in finance and investment. A robust SVR model requires appropriate parameter values, including the kernel function, regularization ”C”, and epsilon parameter ”ϵ”. The SVM model requires a specific data format, and a common technique used in this case is data standardization. This technique involves scaling the data to make the mean of the observed values 0 and the standard deviation 1. This can be achieved using the StandardScaler function, which is represented by Eq. (1).

$$z = \frac{(x - mean)}{standard\_deviation} \quad (1)$$

3.6 Fitting the LSTM Data

LSTM networks are a type of recurrent neural network (RNN). LSTM can work in different areas and problems, one of which is the prediction problems underpinning our study. We applied two types of LSTM models: with and without dropout techniques. Dropout is one of the most types of regularization techniques used to avoid overfitting within model training. The model has been fitted with 100 epochs with a batch size of 1; to take the whole dataset. Also, we have used the Rectified Linear Unit (ReLU) activation function, the "mean squared error" as the loss function, and "adam" as the optimization algorithm. The LSTM model requires a specific data format, and a common technique used in this case is data normalization. This involves rescaling the data from its original format so that all values are within the range of 0 and 1. To achieve this, we used the MinMaxScaler function, as shown in Eq. (2).

$$z = \frac{(x - min)}{max/_{min}} = \frac{\min(x - \min)}{max} \quad (2)$$

3.7 Splitting

One of the important steps in the preprocessing phase is splitting the data into training and testing sets. There are many functions for splitting data. In this research, we used the TimeSeriesSplit function. It provides training and test indices to split the data that is observed at time intervals. The TimeSeriesSplit function splits the train and test data into many split numbers. In each split, the train data will be more than the test set. Our data has been split 10 times, and each time the training data is more than the test data. It ensures that the test datasets are younger or later than the train datasets. The final number of records for the train and test set are 36312 and 3631, respectively.

3.8 Evaluation Metrics

We used three evaluation metrics, Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE).

MSE measures the amount of error in the model. It is the average of the summation of the square difference between the predicted value (Y) and the actual value (X). This function can be found using Eq. (3) [31].

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(X_i - Y_i)^2 \quad (3)$$

RMSE is the standard deviation of the prediction error. It is the average of the summation of the square difference between the predicted value (Y) and the actual value (X) under the square root. This function can be found using Eq. (4) [31].

$$RMSE = \sqrt{MSE} \quad (4)$$

MAE is the average of all absolute errors. Absolute error is the difference between the predicted value (Y) and the actual value (X). This function can be found using Eq. (5) [31].

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|X_i - Y_i| \quad (5)$$

All the steps explained above presented in Fig. 4. This diagram displays the sequence of events have been done on the data that used in our study from data preprocessing into model evaluation.

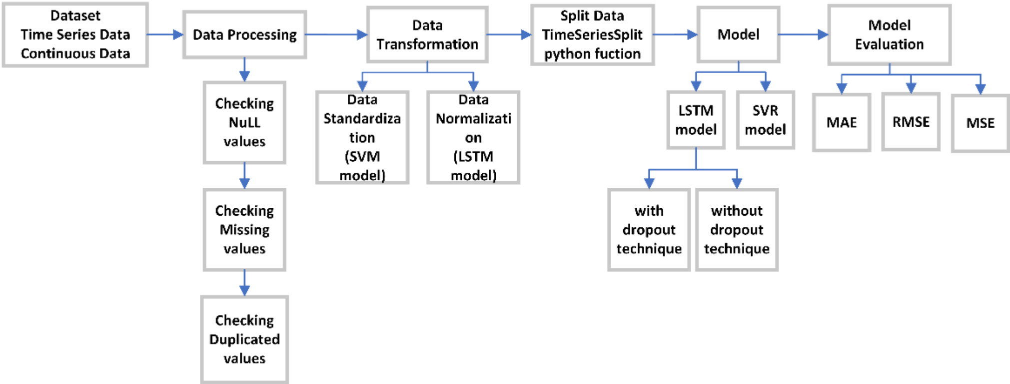

Fig 4. Study sequence events

4. SVM Experiments

In order to identify the optimal hyperparameters for the Support Vector Regression (SVR) model. Various experiments were conducted to determine the optimal statistical model with the lowest error. The values of ”C” and ”ϵ” were investigated by experimenting with different values of ”C” and ”ϵ” while the kernel function remained fixed at the RBF function. Based on the evaluation metrics presented in Table 5, it was found that, ”C” = 100 and,”ϵ” = 0.0005 provided the smallest error. In order to confirm that these values were indeed optimal, additional experiments were conducted with different values of ”C” while ”ϵ” was fixed at 0.0005. The results of these experiments are presented in Tables 5 and 6.

5. LSTM Experiments

Two experiments were conducted in this study, one with and one without the dropout technique, using different numbers of neurons, namely 4, 45, 50, 55, 123, 128, and 133. The LSTM model with dropout technique outperformed the LSTM without dropout, as evidenced by the lower MSE, RMSE, and MAE metrics with values of 0.000124763, 0.011169727, and 0.009058733, respectively (Table 7 and 8). A comparison between the SVR and LSTM models is presented in Table 9 and Fig. 5. The LSTM model produced the best results, with an MSE of 0.000124763, an RMSE of 0.011169727, and an MAE of 0.009058733.

Table 5. Epsilon Parameter Experiments

Iteration	Kernel	C	Epsilon	MSE	RMSE	MAE
1	Rbf	1	0.002	0.000998061	0.031592099	0.021596322
			0.005	0.000998061	0.031592099	0.021596322
			0.0002	0.000993428	0.03151869	0.021474721
			0.0005	0.000992489	0.031503797	0.021457338
2	Rbf	5	0.002	0.000990073	0.031465433	0.021408276
			0.005	0.001001277	0.031642968	0.021941718
			0.0002	0.000991798	0.031492826	0.021335927
			0.0005	0.000992204	0.031499275	0.021359434
3	Rbf	10	0.002	0.00098526	0.031388851	0.02131525
			0.005	0.000996085	0.03156081	0.021886207
			0.0002			

			0.0005	0.00098411	0.031370525	0.021169273
				0.000985101	0.031386322	0.0212074
4	Rbf	100	0.002	0.000964943	0.031063525	0.020809718
			0.005	0.000966381	0.031086668	0.021290104
			0.0002	0.000967086	0.031098014	0.020692475
			0.0005	0.000964316	0.031053438	0.020681691

Table 6. Values of Kernel, C, and epsilon parameters with their MSE, RMSE, and MAE metrics

Iteration	Kernel	C	Epsilon	MSE	RMSE	MAE
1	Rbf	1	0.0005	0.000992489	0.031503797	0.021457338
2	Rbf	5	0.0005	0.000992204	0.031499275	0.021359434
3	Rbf	10	0.0005	0.000985101	0.031386322	0.0212074
4	Rbf	100	0.0005	0.000964316	0.031053438	0.020681691
5	Rbf	0.1	0.0005	0.000995846	0.031557025	0.02154055
6	Rbf	0.5	0.0005	0.000996214	0.031562865	0.021531734
7	Rbf	0.01	0.0005	0.00100925	0.031768703	0.02163933
8	Rbf	0.05	0.0005	0.000997686	0.031586164	0.02156518
9	Rbf	0.005	0.0005	0.001028449	0.032069435	0.022483954
10	Rbf	0.001	0.0005	0.001038608	0.032227437	0.023295853
11	Rbf	0.0001	0.0005	0.001219073	0.03491523	0.02861767
12	Rbf	0.0005	0.0005	0.001054227	0.032468858	0.024017888

Table 7. LSTM without dropout technique

Iteration	units	Dense	Epoch	Batch Size	MSE	RMSE	MAE
1	4	1	100	1	0.000371398	0.019271689	0.017944673
2	45	1	100	1	0.000435148	0.020860211	0.017465161
3	50	1	100	1	0.00038451	0.01960892	0.016233417
4	55	1	100	1	0.000383561	0.019584722	0.016097014
5	123	1	100	1	0.00035296	0.018787234	0.015499975
6	128	1	100	1	0.000278915	0.016700739	0.01404591
7	133	1	100	1	0.00034101	0.018466464	0.015146258

Table 8. LSTM with dropout technique

Iteration	units	Dense	Epoch	Batch size	Drop out	MSE	RMSE	MAE
1	4	1	100	1	0.2	0.000350706	0.018727141	0.016360837
2	45	1	100	1	0.2	0.000167408	0.012938639	0.010304672
3	50	1	100	1	0.2	0.000124763	0.011169727	0.009058733
4	55	1	100	1	0.2	0.000291197	0.017064482	0.013871041
5	123	1	100	1	0.2	0.000274765	0.016576023	0.013768943
6	128	1	100	1	0.2	0.000183983	0.013564045	0.010960331
7	133	1	100	1	0.2	0.000161595	0.012712005	0.010213781

Table 9. Summary of results for the SVR and LSTM models

Model	MSE	RMSE	MAE
LSTM	0.000124763	0.011169727	0.009058733
SVR	0.000964316	0.031053438	0.020681691

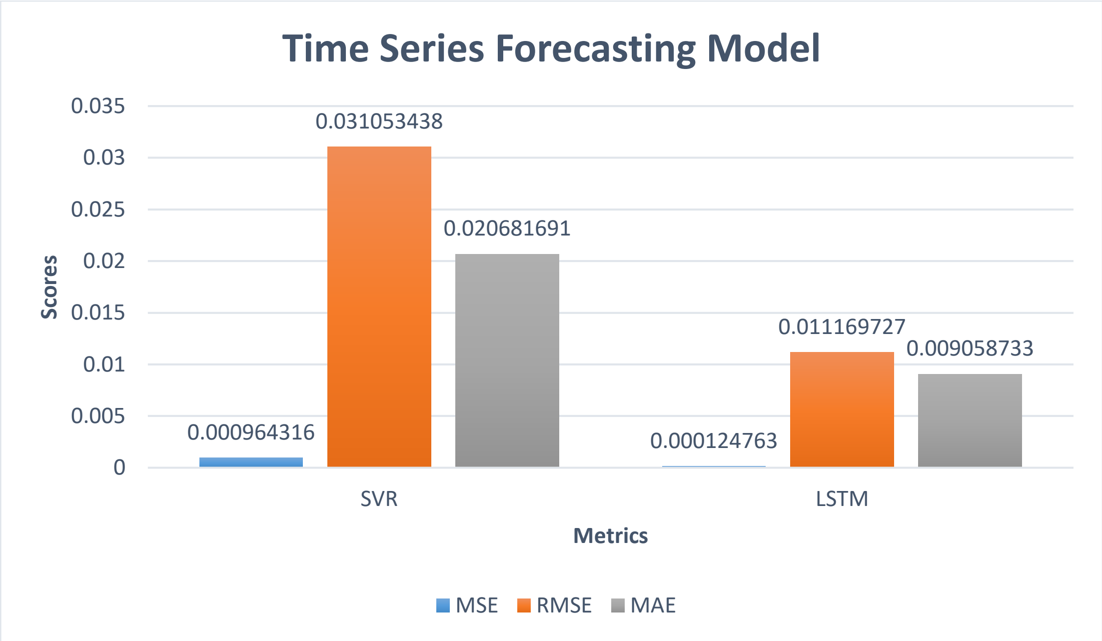

Fig 5. Time Series Forecasting Model

6. Results and Discussion

This study aimed to determine the optimal time for trading and investing in stocks by identifying the most accurate time series forecasting model. To this end, two machine learning and deep learning models: Support Vector Regression (SVR) and Long Short-Term Memory (LSTM), respectively were used.

In SVR experiments, various experiments were conducted to determine the optimal statistical model with the lowest error. The values of ”C” and ”ϵ” were investigated by experimenting with different values of ”C” and ”ϵ” while the kernel function remained fixed at the RBF function.

On the other side, two experiments were conducted in this study using the LSTM, with and without the dropout technique, using different numbers of neurons, namely 4, 45, 50, 55, 123, 128, and 133. The LSTM model with dropout technique outperformed the LSTM without dropout.

Our results indicated that the LSTM model outperformed the SVR model, achieving the lowest MSE, RMSE, and MAE values of 0.000124763, 0.011169727, and 0.009058733, respectively. These evaluation metrics are crucial in assessing the accuracy of a model. MSE measures the level of error in a model, with lower values indicating a better fit to the dataset. RMSE calculates the average error between predicted and actual values, while MAE provides the average difference between predicted and actual values (absolute error).

The output values from the model should be the best time for trading. These values showed the best time for trading based on the data trained in the model. So, you should take these values into consideration when trading in the future. At the end, in the financial world, nothing is inevitable, every fraction of a second can change many things. So what we have to do is close if not hit. Some output values from the LSTM model explained in table 10.

Table 10. Sample of the best time for trading based on the LSTM model

Date	Time
2022-07-27	11:18:24
2022-08-14	07:32:16
2022-06-27	11:09:52
2022-09-30	09:40:16
2022-11-27	11:22:40
2022-07-30	06:06:56

7. Conclusion and Future Work

Knowing the correct time for trading is essential for investors and traders who want to increase their profits and decrease their losses. Identifying the optimal time to buy or sell stocks is a critical factor in successful trading.

In this research, we tried to determine the optimal time for trading stocks and options by building a time series forecasting model. Historical data for a variety of Japanese stocks and options was employed, and the dataset was prepared for use in the models. Two models were employed: Support Vector Regression (SVR) as a type of Support Vector Machine (SVM) for machine learning, and Long-Short Term Memory (LSTM) for deep learning. The study showed the superiority of deep learning on machine learning model. LSTM model outperformed the SVR model by providing the lowest values for the evaluation metrics used in this study. Also, LSTM model with dropout technique outperformed the LSTM without dropout technique.

In future work, we plan to augment the dataset size, which can improve model efficiency and lead to better results. Additionally, we will perform further experiments to optimize the parameters for both models used in this research.

Acknowledgment

This research work was carried out as part of the requirements for the completion of the Master's program at Jordan University of Science and Technology (JUST). The author would like to acknowledge the support and guidance provided by the faculty members and the research team at JUST.

References

[1] A. HAYES. What is a time series and how is it used to analyze data? [Online]. Available: [4.2023] What Is a Time Series and How Is It Used to Analyze Data? (investopedia.com)

[2] M. K. Okasha. (2014). "Using support vector machines in financial time series forecasting," International Journal of Statistics and Applications, vol. 4, no. 1, pp. 28–39.

[3] R. J. Hyndman and G. Athanasopoulos, Forecasting: principles and practice. OTexts, 2018. [Online]. Available: accessed 2023-4-24

[4] N. Kumar and S. Susan. (2020). "Covid-19 pandemic prediction using time series forecasting models," In 11th international conference on computing, communication and networking technologies (ICCCNT). IEEE, 2020, pp. 1–7.

[5] E. Pintelas, I. E. Livieris, S. Stavroyiannis, T. Kotsilieris, and P. Pintelas. (2020). "Investigating the problem of cryptocurrency price prediction: a deep learning approach," in Artificial Intelligence Applications and Innovations: 16th IFIP WG 12.5 International Conference, AIAI 2020, Neos Marmaras, Greece, June 5–7, 2020, Proceedings, Part II 16. Springer, 2020, pp. 99–110.

[6] M. Lam. (2004). "Neural network techniques for financial performance prediction: integrating fundamental and technical analysis," Decision support systems, vol. 37, no. 4, pp. 567–581.

[7] R. A. A. Viadinugroho and D. Rosadi. (2021). "Long short-term memory neural network model for time series forecasting: case study of forecasting ihsg during covid-19 outbreak," in Journal of Physics: Conference Series, vol. 1863, no. 1. IOP Publishing, p. 012016.

[8] F. E. Tay and L. Cao. (2001). "Application of support vector machines in financial time series forecasting," omega, vol. 29, no. 4, pp. 309–317.

[9] H. A. H. Al-Tamimi, A. A. Alwan, and A. Abdel Rahman. (2011). "Factors affecting stock prices in the uae financial markets," Journal of transnational management, vol. 16, no. 1, pp. 3–19.

[10] I. Yanovitzky and A. VanLear. (2008). "Time series analysis: Traditional and contemporary approaches".

[11] J. V. Hansen and R. D. Nelson. (1997). "Neural networks and traditional time series methods: a synergistic combination in state economic forecasts," IEEE transactions on Neural Networks, vol. 8, no. 4, pp. 863–873.

[12] Y. Tang, Z. Song, Y. Zhu, H. Yuan, M. Hou, J. Ji, C. Tang, and J. Li. (2022). "A survey on machine learning models for financial time series forecasting," Neurocomputing, vol. 512, pp. 363–380.

[13] V. Derbentsev, A. Matviychuk, N. Datsenko, V. Bezkorovainyi, and A. Azaryan. (2020). "Machine learning approaches for financial time series forecasting." CEUR Workshop Proceedings.

[14] M. Cai, M. Pipattanasomporn, and S. Rahman. (2019). "Day-ahead building-level load forecasts using deep learning vs. traditional time-series techniques," Applied energy, vol. 236, pp. 1078–1088.

[15] J.-F. Chen, W.-L. Chen, C.-P. Huang, S.-H. Huang, and A.-P. Chen. (2016). "Financial time-series data analysis using deep convolutional neural networks," In 7th International conference on cloud computing and big data (CCBD). IEEE, 2016, pp. 87–92.

[16] O. B. Sezer, M. U. Gudelek, and A. M. Ozbayoglu. (2020). "Financial time series forecasting with deep learning: A systematic literature review: 2005–2019," Applied soft computing, vol. 90, p. 106181.

[17] A. Dingli and K. S. Fournier. (2017). "Financial time series forecasting-a deep learning approach," International Journal of Machine Learning and Computing, vol. 7, no. 5, pp. 118–122.

[18] A. Barbulescu and E. Bautu. (2012). "A hybrid approach for modeling financial time series." Int. Arab J. Inf. Technol., vol. 9, no. 4, pp. 327–335.

[19] K.-j. Kim. (2003). "Financial time series forecasting using support vector machines," Neurocomputing, vol. 55, no. 1-2, pp. 307–319.

[20] Y. Ensafi, S. H. Amin, G. Zhang, and B. Shah. (2022). "Time-series forecasting of seasonal items sales using machine learning–a comparative analysis," International Journal of Information Management Data Insights, vol. 2, no. 1, p. 100058.

[21] W.-H. Chen, J.-Y. Shih, and S. Wu. (2006). "Comparison of support-vector machines and back propagation neural networks in forecasting the six major asian stock markets," International Journal of Electronic Finance, vol. 1, no. 1, pp. 49–67.

[22] A. Dutta, G. Pooja, N. Jain, R. R. Panda, and N. K. Nagwani. (2021). "A hybrid deep learning approach for stock price prediction," in Machine Learning for Predictive Analysis: Proceedings of ICTIS 2020. Springer, pp. 1–10.

[23] C. F. Tsai and S. P. Wang. (2009). "Stock price forecasting by hybrid machine learning techniques," in Proceedings of the international multiconference of engineers and computer scientists, vol. 1, no. 755, p. 60.

[24] S. Siami-Namini, N. Tavakoli, and A. S. Namin. (2018). "A comparison of arima and lstm in forecasting time series," in 2018 17th IEEE international conference on machine learning and applications (ICMLA). IEEE, pp. 1394–1401.

[25] S. Mehtab and J. Sen. (2020). "Stock price prediction using convolutional neural networks on a multivariate timeseries," arXiv preprint arXiv:2001.09769.

[26] A. Gupta and H. Nain. (2021). "Bitcoin price prediction using time series analysis and machine learning techniques," in Machine Learning for Predictive Analysis: Proceedings of ICTIS 2020. Springer, pp. 551–560.

[27] C. Tandon, S. Revankar, and S. S. Parihar. (2021). "How can we predict the impact of the social media messages on the value of cryptocurrency? insights from big data analytics," International Journal of Information Management Data Insights, vol. 1, no. 2, p. 100035.

[28] M. Iqbal, M. Iqbal, F. Jaskani, K. Iqbal, and A. Hassan. (2021). "Time-series prediction of cryptocurrency market using machine learning techniques," EAI Endorsed Transactions on Creative Technologies, vol. 8, no. 28.

[29] E. Sin and L. Wang. (2017). "Bitcoin price prediction using ensembles of neural networks," in 2017 13th International conference on natural computation, fuzzy systems and knowledge discovery (ICNC-FSKD). IEEE, pp. 666–671.

[30] J. Brownlee, "How to check if time series data is stationary with python".[online] machine learning mastery," 2020.

[31] A. Chugh, "Mae, mse, rmse, coefficient of determination, adjusted r squared—which metric is better," Retrieved December, vol. 8, p. 2022, 2020.

Notes on contributors

Dr. Qusai Abuein an associate professor in the CIS / JUST. He received his PhD and master's degrees Japan. His research interests include web social analysis, web analytics, data analysis, and information retrieval. He can be contacted at: qabuein@just.edu.jo, ORCID account: https://orcid.org/0009-0002-4323-4301, copus accpumt: https://www.scopus.com/authid/detail.uri?authorId=6506320701, google scholar accoumt: https://scholar.google.com/citations?user=s0l6AgwAAAAJ&hl=en

Dr. Mohammed Q Shatnawi

Dr. Shatnawi completed his master and Ph.D. from the George Washington University/ DC in 2007. He is currently working in the Jordan University of Science and Technology. His research interests are in deep learning, information and data security, big data analysis, project management and NLP. He can be contacted at: mshatnawi@just.edu.jo. Scopus account: https://www.scopus.com/authid/detail.uri?authorId=24723220900 , ORCID account: https://orcid.org/0000-0002-5560-3592, google scholar account: https://scholar.google.com/citations?user=IZ3KGVsAAAAJ&hl=en

Emran Y. Aljawarneh is a researcher at the Department of Computer Information Systems, Jordan University of Science and Technology (JUST), Irbid, Jordan. Aljawarneh received his MSc. from JUST in 2022. His research interest includes Time Series analysis and data analysis. Aljawarneh has published several research articles in international journals of Information Technology.

Dr. Ahmad M. Manasrah is an associate professor of computer sciences at Higher Colleges of Technology (HCT), Sharjah, UAE since 2018, prior to joining HCT, Dr. Manasrah, was an associate professor at the faculty of information system and computer science at Yarmouk University, Jordan, since 2011. Dr. Manasrah has published over 40 international journal and 14 research papers including proceedings and 4 chapters of book, and 3 international patents in the field of network monitoring and security. Dr. Manasrah received his PhD. from University Sains Malaysia (USM) in 2009. His research interest includes advanced Internet security and monitoring.